DOCUMENT RESUME

ED 314 429 TM 014 147

AUTHOR Williams, Paul L.
TITLE Using Customized Standardized Tests. ERIC Digest.
INSTITUTION American Institutes for Research, Washington, DC.; ERIC Clearinghouse on Tests, Measurement, and Evaluation, Washington, DC.
SPONS AGENCY Office of Educational Research and Improvement (ED), Washington, DC.
REPORT NO EDO-TM-89-3
PUB DATE Dec 89
CONTRACT RI88062003
NOTE 3p.; For document from which this Digest was extracted, see TM 014 144.
PUB TYPE Information Analyses - ERIC Information Analysis Products (071)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Achievement Tests; *Criterion Referenced Tests; Elementary Secondary Education; *Norm Referenced Tests; *Standardized Tests; State Programs; *Test Construction; Test Format; Testing Problems; Testing Programs; Test Interpretation; Test Norms; Test Use
IDENTIFIERS ERIC Digests; Fact Sheets; *Test Customization

ABSTRACT

Both norm-referenced and criterion-referenced interpretations of student, school, district, and state testing data can be improved by customizing the traditional norm-referenced test. Improvements that can be made include: (1) increasing the relevance of the test to the curriculum and (2) having greater confidence in the national comparative information. Testing time can also be reduced. For the past several years, Texas has used a model state criterion-referenced test that was statistically equated to a nationally norm-referenced test. In spite of the reduced testing time and greater relevance to the curriculum of this approach, Texas will leave this model in 1990 for one that may be more successful in producing scores that approach norm validity. A second model of a custom-made test is one in which state or district-developed criterion-referenced items are combined with a complete norm-referenced test. Norm invalidity may still be a problem if norm-referenced scores are inflated by instruction targeted at the objectives. In Tennessee, a third model of a customized test remedies the shortcomings of the first two models and provides objective scores for instructional planning and norm-referenced scores for national comparisons. Customized norm-referenced tests offer a viable alternative to both norm-referenced and criterion-referenced tests. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED314429

ERIC Clearinghouse on Tests, Measurement, and Evaluation Digest EDO-TM-89-3 December 1989

Using customized standardized tests

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
✓ This document has been reproduced as received from the person or organization originating it.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Paul L. Williams, CTB/McGraw-Hill

Over the next several years it is likely that you'll see a subtle but important change in the nature of standardized tests that are administered as part of your state and district testing programs. This change results from a desire to improve both the norm- and criterion-referenced interpretations of student, school, district, and state testing data. These interpretations can be improved by customizing the traditional norm-referenced test.

Norm-referenced tests are designed to give you both normative and objective information. Normative information may take the form of scale scores, percentile ranks, grade equivalents, normal curve equivalents, and stanines. Objective performance is usually reported as a percentage master score based on the objectives included on the norm-referenced test.

Normative scores allow you to compare individuals and groups with national performance levels, and objective scores allow you to make comparisons relative to specific objectives. Together, these scores allow you to plan programs for your school and district and instruction for individual students.

When used correctly, this information is invaluable for school administrators. However, several improvements can be made so that you can make even better programmatic and individual plans, such as

- reducing testing time,
- increasing the relevance of the test to the curriculum, and
- having greater confidence in the national comparative information.

These improvements are the goals of custom-made norm-referenced tests.

Several models for constructing custom-made norm-reference tests have been attempted, with some degree of success. A discussion of three models follows.

A model used in Texas

For the last few years, Texas has used a model state criterion-referenced test, which was statistically equated to a nationally normed norm-referenced test. Texas now administers the criterion-referenced test instead of the norm-referenced test and both norm-referenced and criterion-referenced scores are produced.

The advantages of this approach are reduced testing time and greater relevance to the Texas curriculum than could be obtained from using the norm-referenced test alone.

However, this approach has several disadvantages:

- Equating these two different tests will result in inaccurate norm-referenced scores because of differences in test difficulty and content between the norm-referenced and criterion-referenced tests. Criterion-referenced scores are unaffected by the equating.
- Instruction focused on the curriculum will likely increase both the criterion-referenced scores and, as a result, the equated norm-referenced scores. Although score increases on the criterion-referenced portion of the test may accurately reflect student learning in these restricted domains, this is not the case for the much broader norm-referenced domains.

This is because instruction has been effectively focused on only a portion of the traits measured by norm-referenced tests, thus producing higher equated norm-referenced scores than would be expected if the original norm-referenced test or a proper sample of items from that test were administered.

When this distortion happens, the norm-referenced scores produced from this model are called norm-invalid. That is, the customized test does not accurately reproduce the normative scores that would have resulted had the entire norm-referenced test been administered.

For a custom-made norm-referenced test to be fair, the scores must be norm-valid (Yen, Green, and Burket, 1987). Texas will leave this model in 1990 in favor of one that may be more successful in producing scores that approach norm-validity.

A second model

A second model of a custom-made test is one in which state- or district-developed criterion-referenced items are combined with a complete norm-referenced test. Norm-referenced scores are generated from the complete norm-referenced test, while objective information is derived from a combination of norm-referenced and locally developed items.

This type of test reduces testing time because only one customized test is administered instead of both a

TM014141

norm-referenced and a criterion-referenced test. However, as with the Texas model that we discussed, norm invalidity may be a problem.

If instruction is carefully targeted at the objectives and a subset of the norm-referenced test items is used for reporting achievement by objective, then norm-invalidity could result because instruction influences only a portion of the trait measured by the norm-referenced test. In this case, the norm-referenced scores could be inflated by the targeted instruction, thus rendering them invalid.

A model used in Tennessee

Another model of a customized test was recently adopted by the State of Tennessee. The Tennessee model remedies the shortcomings of the first two models that we described. This model uses approximately 40 items instead of a full-length test of 80 to 110 items for its norm-referenced module and a criterion-referenced module of state-developed items.

The norm-referenced module was specifically created so that it has proper statistical characteristics of reliability, adequate floors and ceilings, and articulation across test levels. Tennessee will use multiple test forms.

Items used for the norm-referenced portion are not intended to be used for objective scores, and the criterion-referenced items are not used as part of the norm-referenced scores.

Effective instruction targeted toward the state objectives will demonstrate student attainment of the state's objectives, and the norm-referenced portion will provide norm-valid scores. Thus, the Tennessee model reduces testing time and requires only one testing period rather than two. The objective scores will be useful for instructional planning and the norm-referenced scores can be used with confidence for national comparisons.

A note about norm-validity

As a school administrator, you should be concerned about the norm-validity of your district's test scores. During times of increased school, district, state, and national achievement (as we see now), critics may be quick to question the validity of your test results. Critics may point out that teachers are too familiar with the test items, that they teach actual test items, or that the scores may not reflect true changes in achievement. Williams (1988) and Koretz (1988a, 1988b) have both presented a distinction between changes in test scores and changes in achievement.

Changes in test scores may result from a variety of instructional and administrative interventions, but changes in test scores may not reflect actual changes in achievement. Special coaching, inappropriate test preparation materials and methods, and narrowly targeted instruction may all increase test scores, but they do not necessarily lead to sustained and abiding increases in achievement.

Just as instruction must support test score changes that are not spurious, i.e. produce true growth, test instruments must be designed and implemented so that if score increases occur, they represent a true change in achievement and are not the result of an inadequately designed customized testing program.

Unless a customized norm-referenced test produces norm-valid scores, you cannot provide test results that reflect true changes in achievement. Even with an optimally designed customized test, abuses can still result. But without a properly designed customized norm-referenced test, you cannot demonstrate that achievement, rather than just test scores, has improved.

Administrators at all levels must be able to tell the difference between norm-valid tests that allow actual achievement to be demonstrated and norm-invalid ones. When norm-valid test are used, you can report the test results with confidence.

If you have confidence in the test's quality, then test scores will accurately reflect meaningful changes in student achievement. Thus, you will be able to determine the effectiveness of your instructional program.

If you have a norm-valid test, you can show your constituents that changes in the test scores are real. When these changes represent increases, your community and staff can be satisfied the instructional program works in the areas the test measures. If the score changes represent a decrease, then the test results can help you identify areas that need additional instructional effort. In either case, the students win because instructional support is forthcoming.

Customized norm-referenced tests offer a viable alternative to both norm-referenced and criterion-referenced tests. One test, instead of two, is all that needs to be administered. Disruption in the schools is reduced, testing time is reduced, and instructional time is maximized. Alternate forms of customized norm-referenced tests can be used, minimizing criticisms of test familiarity and inappropriate test preparation activities. Teachers will be more likely to teach the complete curriculum, and increased achievement, rather than just increased scores, can result.

References

Koretz, D. (Summer, 1988a). "Arriving in Lake Wobegon: Are Standardized Tests Exaggerating Achievement and Distorting Instruction?" *American Educator*, 8-15, 46-52.

Koretz, D. (1988b). Panel presentation, ECS Large-Scale Assessment Conference, Boulder, Colorado.

Williams, P.L. (1988). Panel presentation, ECS Large-Scale Assessment Conference, Boulder, Colorado.

Yen, W. M., Green, D. R., and Burket, G. R. (1987). "Valid Normative Information from Customized Achievement Tests." *Educational Measurement: Issues and Practice*, 6, 7-13.

This publication was prepared with funding from the Office of Educational Research and Improvement (OERI), U.S. Department of Education, under contract R-88-062003. The opinions expressed in this report do not necessarily reflect the position or policy of OERI or the Department of Education. **Permission is granted to copy and distribute this ERIC/TM digest.**